WHAT IS CLAIMED IS:

1. A method for exchanging data between compute nodes of a computer system comprising:

a plurality of compute nodes interconnected by an inter-node communication network, each of the compute nodes having an independent address space and comprising:

a local packetized interconnect,

a network interface coupled to the local packetized interconnect and the inter-node communication network,

at least one data processor coupled to the local packetized interconnect; and,

a memory system coupled to the local packetized interconnect;

the method comprising tunneling data from the sending compute node to the receiving compute node by:

placing a local packetized interconnect packet on the local packetized interconnect of the sending compute node;

receiving the local packetized interconnect packet at the network interface of the sending compute node;

encapsulating the local packetized interconnect packet in an inter-node communication network packet addressed to the receiving compute node;

dispatching the inter-node communication network packet to the receiving compute node by way of the inter-node communication network;

receiving the inter-node communication network packet at the network interface of the receiving compute node;

extracting the local packetized interconnect packet from the inter-node communication network packet; and,

placing the extracted packet onto the local packetized interconnect of the receiving compute node

wherein, when the local packetized interconnect packet is on the local packetized interconnect of the sending compute node, portions of the local packetized interconnect packet other than any addresses and any check values are substantially the same as the corresponding portions of the local packetized interconnect packet when the local packetized interconnect packet is on the local packetized interconnect of the receiving compute node.

2. A method according to claim 1 comprising associating a first range of addresses in an address space of a sending one of the compute nodes with the network interface of the sending compute node and, at the network interface of the sending compute node associating the first range of addresses with the receiving compute node wherein the method comprises determining that the local packetized interconnect packet is associated with an address in the first range of addresses upon receiving the local packetized interconnect packet at the network interface of the sending compute node.

3. A method according to claim 2 wherein the method comprises performing an address translation on the local packetized

interconnect packet after receiving the packet at the network interface of the sending compute node and prior to placing the extracted packet onto the local packetized interconnect of the receiving compute node.

4. A method according to claim 3 wherein performing the address translation comprises editing the local packetized interconnect packet by changing an address to which the packet is addressed from an address in the first range of addresses to a corresponding address in an address space of the receiving compute node.

5. A method according to claim 4 wherein the address translation is performed at the network interface of the sending compute node.

6. A method according to claim 5 comprising allocating a region of the memory of the receiving compute node to receive data from the sending compute node, memory locations in the allocated region being addressable by addresses in a second range of addresses corresponding to the first range of addresses, and communicating the second range of addresses from the receiving compute node to the sending compute node prior to tunneling the data from the sending compute node to the receiving compute node.

7. A method according to claim 6 comprising, at the sending compute node, prior to tunneling the data, computing an address transformation between the first and second address ranges and using the address transformation to perform the address translation.

8. A method according to claim 7 wherein the address transformation comprises an address translation table.

9. A method according to claim 6 wherein the first and second ranges of addresses are equal in size.

10. A method according to claim 4 wherein the address translation is performed at the network interface of the receiving compute node.

11. A method according to claim 1 wherein the local packetized interconnect packet comprises a write request packet comprising data to be written to a memory location in the memory system of the receiving compute node corresponding to the address and the method comprises writing the data to the memory location in the receiving compute node.

12. A method according to claim 11 wherein the write request packet comprises a write confirmation request and the method comprises, detecting the write confirmation request at the network interface of the sending compute node, generating a write confirmation packet at the network interface of the sending compute node and dispatching the write confirmation packet on the local packetized interconnect of the sending compute node.

13. A method according to claim 12 comprising, at the network interface of the sending compute node, retaining a copy of the write request packet, maintaining a write completion timer and, if the

write completion timer times out, using the copy of the write request packet to resend the write request packet to the receiving compute node by way of the inter-node communication network.

14. A method according to claim 11 wherein the write request packet comprises a write confirmation request and the method comprises:

detecting the write confirmation request at the memory system of the receiving compute node,

placing on the local packetized interconnect of the receiving compute node a write confirmation packet containing the address of the memory location in the address space of the receiving compute node;

changing the write confirmation packet by altering the contained address of the memory location in the address space of the receiving compute node to a corresponding address in the first range of addresses; and,

subsequently placing the write confirmation packet on the local packetized interconnect of the sending compute node.

15. A method according to claim 14 wherein changing the write confirmation packet is performed at the network interface of the sending compute node.

16. A method according to claim 14 wherein changing the write confirmation packet is performed at the network interface of the receiving compute node.

17. A method according to claim 11 wherein the write request packet comprises a write confirmation request and the method comprises, detecting the write confirmation request at the network interface of the receiving compute node, generating a write confirmation packet at the network interface of the receiving compute node and dispatching the write confirmation packet to the sending compute node.

18. A method according to claim 17 comprising, at the network interface of the receiving compute node, retaining a copy of the write request packet, maintaining a write completion timer and, if the write completion timer times out, using the copy of the write request packet to resend the write request packet to the memory system of the receiving compute node.

19. A method according to claim 1 wherein the local packetized interconnect packet comprises a read request packet.

20. A method according to claim 2 comprising, at the network interface of the sending compute node, associating the first range of addresses with a plurality of receiving compute nodes.

21. A method according to claim 20 comprising maintaining a correspondence between the first range of addresses and address ranges in address spaces of the plurality of receiving compute nodes.

22. A method according to claim 21 wherein the address ranges in the address spaces of the plurality of receiving compute nodes are not all the same.

23. A method according to claim 20 comprising dispatching the inter-node communication network packet to the plurality of receiving compute nodes by way of a multicast feature of the inter-node communication network.

24. A method according to claim 20 comprising, at the sending compute node, encapsulating each of a plurality of copies of the local packetized interconnect packet in an inter-node communication network packet addressed to a different one of the plurality of receiving compute nodes and dispatching each of the inter-node communication network packets to the corresponding receiving compute node by way of the inter-node communication network.

25. A method according to claim 2 comprising placing a plurality of local packetized interconnect packets addressed to one or more addresses in the first range of addresses on the local packetized interconnect of the sending compute node wherein encapsulating the packet in the inter-node communication network packet comprises encapsulating the plurality of local packetized interconnect packets in the same inter-node communication network packet.

26. A method according to claim 25 wherein the plurality of local packetized interconnect packets each comprise header information

and a payload, at least some of the header information is the same for each of the local packetized interconnect packets and encapsulating the plurality of local packetized interconnect packets comprises placing into the inter-node communication network packet the payloads of all of the plurality of local packetized interconnect packets and the header information from fewer than all of the plurality of local packetized interconnect packets.

27. A method according to claim 1 wherein the sending and receiving compute nodes are peers.

28. A method according to claim 27 wherein the sending compute node has substantially the same construction as the receiving compute node.

29. A method according to claim 1 wherein the local packetized interconnect packet comprises an atomic read-modify-write packet.

30. A method according to claim 1 wherein the local packetized interconnects of the compute nodes operate at data rates in excess of 300 MBps.

31. A method according to claim 30 wherein the inter-node communication network is characterized by a link data rate of at least 1 GBps.

32. A method according to claim 10 comprising, at the receiving compute node, altering a correspondence of the first range of addresses to addresses in the address space of the receiving compute node by changing the second range of addresses to a third range of addresses.

33. A method according to claim 1 wherein the packet placed on the local packetized interconnect of the sending compute node is a read response packet and the method comprises converting the read response packet into a write request packet.

34. A method according to claim 33 comprising issuing a read request directed to a location in the memory system of the sending compute node wherein, placing the packet on the local packetized interconnect of the sending compute node is performed by the memory system of the sending compute node in response to the read request.

35. A method according to claim 1 comprising allocating a memory ID to a range of addresses in an address space of the receiving compute node and associating a corresponding first range of addresses in an address space of the sending compute node with the memory ID and the receiving compute node; at the sending compute node, determining an offset from an address associated with the local packetized interconnect packet and the first range of addresses and passing the memory ID and the offset in the inter-node communication network packet to the receiving compute node.

36. A method according to claim 35 comprising, at the network interface of the receiving compute node, determining from the offset and memory ID a destination address in the range of addresses in the address space of the receiving compute node wherein placing the extracted packet onto the local packetized interconnect of the receiving compute node comprises addressing the extracted packet to the destination address.

37. A method according to claim 2 comprising establishing a reciprocal pathway for tunneling data packets from the receiving compute node to the sending compute node.

38. A method according to claim 1 comprising, at the network interface of the receiving compute node, receiving a request from the data processor of the receiving compute node that data be obtained from the sending compute node and, in response thereto, generating a local packetized interconnect read request packet, encapsulating the read request packet, and forwarding the read request packet to the sending compute node by way of the inter-node communication network wherein the local packetized interconnect packet is a read response packet generated in response to the read request packet.

39. A method performed in a multiprocessor computer system comprising:

a plurality of compute nodes interconnected by an inter-node communication network, each of the compute nodes having an independent address space and comprising:

a local packetized interconnect,

a network interface coupled to the local packetized interconnect and the inter-node communication network,

at least one data processor coupled to the local packetized interconnect; and,

a memory system coupled to the local packetized interconnect;

the method comprising:

at the network interface of a sending one of the compute nodes maintaining an association between a first range of addresses in an address space of the sending compute node and a receiving one of the compute nodes;

receiving on the local packetized interconnect of the sending compute node packets addressed to addresses in the first range of addresses;

in response to determining that the packets are addressed to addresses in the first range of addresses, encapsulating the packets in inter-node communication network packets addressed to the receiving compute node; and,

dispatching the inter-node communication network packets to the receiving compute node by way of the inter-node communication network.

40. A method according to claim 39 comprising:

receiving the inter-node communication network packets at the network interface of the receiving compute node;

extracting the local packetized interconnect packets from the inter-node communication network packets; and,

placing the extracted packets onto the local packetized interconnect of the receiving compute node.

41. A method according to claim 40 comprising performing an address translation on the local packetized interconnect packet after receiving the packet at the network interface of the sending compute node and prior to placing the extracted packet onto the local packetized interconnect of the receiving compute node.

42. A method according to claim 41 wherein performing the address translation comprises editing the local packetized interconnect packet by changing an address to which the packet is addressed from an address in the first range of addresses to a corresponding address in an address space of the receiving compute node.

43. A multiprocessor computer system comprising:

a plurality of compute nodes interconnected by an inter-node communication network, each of the compute nodes having an independent address space and comprising:

a local packetized interconnect,

a network interface coupled to the local packetized interconnect and the inter-node communication network,

at least one data processor coupled to the local packetized interconnect; and,

a memory system coupled to the local packetized interconnect;

wherein the network interface of a sending one of the compute nodes maintains an association between a first range of addresses in an address space of the sending compute node with a receiving one of the compute nodes and the network interface of the sending compute node is configured to: receive on the local packetized interconnect packets addressed to an address in the first range of addresses; encapsulate the packets in inter-node communication network packets addressed to the receiving compute node; and, dispatching the inter-node communication network packets to the receiving compute node by way of the inter-node communication network.

44. A multiprocessor computer system according to claim 43 wherein the network interface of the sending compute node includes an address translation facility operative to translate addresses of the packets into corresponding addresses in an address space of the receiving compute node prior to dispatching the inter-node communication network packets.

45. A network interface for use in a compute node of a multiprocessor computer system, the network interface comprising a facility for maintaining associations between one or more ranges of addresses in an address space of the compute node and corresponding other compute nodes and being constructed to: receive packets on a local packetized interconnect addressed to addresses in the one or more

ranges of addresses; encapsulate the packets in inter-node communication network packets addressed to the corresponding receiving compute nodes; and, dispatch the inter-node communication network packets to the corresponding receiving compute node by way of the inter-node communication network.

46. A network interface according to claim 45 comprising an address translation facility configured to edit the local packetized interconnect packets by changing addresses to which the packets are addressed from addresses in the one or more ranges of addresses to corresponding addresses in the address spaces of the corresponding receiving compute nodes prior to dispatching the inter-node communication network packets which encapsulate the local packetized interconnect packets.